Auto-sensing, Auto-negotiation, and Duplexing at JPL

September 12, 2001

Claudia de Luna, claudia.de.luna@jpl.nasa.gov
JPLNet

Introduction

In instances where a user experiences network performance degradation, particularly when it immediately follows a move to a new network connection, the installation of a new network interface card, or an upgrade to the switches servicing the local network, it is often a case of faulty negotiation of operating parameters between the system network interface card and the network equipment. Users may hear the terms “auto-sensing”, “auto-negotiation”, or “duplexing mismatch” to describe the problem.

This problem is not unique to JPL nor is it unique to the equipment in use to support the JPL network. Any enterprise providing 10/100Mb/s Ethernet links to its users, irrespective of equipment vendor or model, faces this problem.

This paper will attempt to describe the auto-negotiation specification and the problems with it with particular emphasis on the networking environment at JPL.

Background

During the development of the 100Base-T standard for 100Mb/s operation, commonly referred to as "Fast Ethernet", the IEEE (802.3u) working group recognized that for this standard to be successful and to "interoperate" with the large existing installed base of 10Base-T Ethernet (10Mb/s) some method by which devices could identify the capabilities of other network components to which they were directly connected and communicate their own capabilities to those components would be essential. This capability would allow two link partners to negotiate to the highest common denominator and thus provide the best possible service over that link. There were several proposals and IEEE selected a specification known as "NWAY" which is defined by Bill Bunch of National Semiconductor in his paper An Introduction to Auto-Negotiation.

Auto-negotiation attempts to negotiate 2 parameters, link speed and link duplexing, between 2 connected devices. The speed choices for Ethernet (10Base-T) and Fast Ethernet (100Base-T) are 10Mb/s and 100Mb/s respectively. The duplexing choices which can apply to either Ethernet or Fast Ethernet are full and half. Full-duplex mode separates the transmit and receive paths and dedicates each path its own wire pair. This is the most efficient operation as it avoids collisions and essentially doubles the throughput of the link. Half-duplex mode is the behavior of the original Ethernet specification.

Auto-Negotiation supports:

10Base-T Half Duplex

10Base-T Full Duplex

100Base-TX Half Duplex

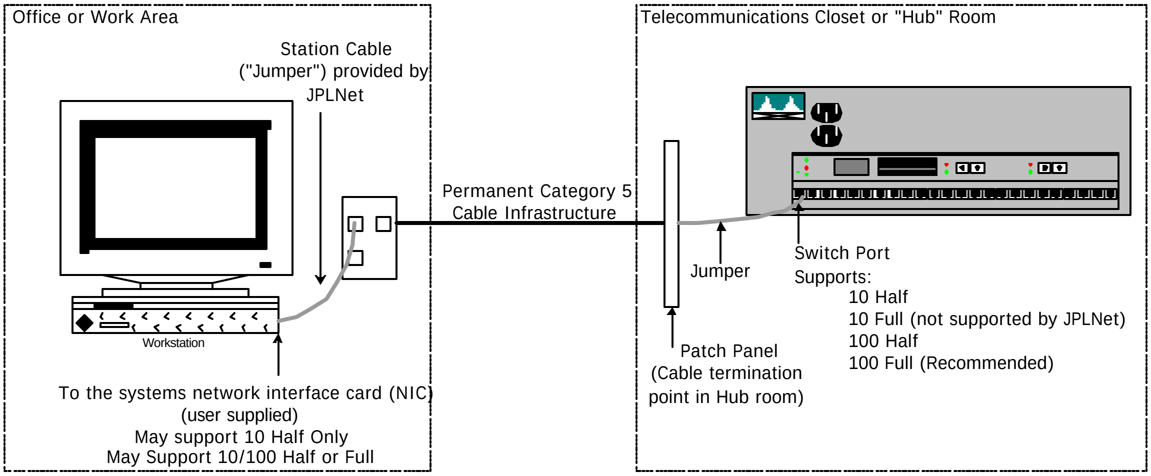
100Base-TX Full Duplex

100Base-T4 (Not supported at JPL. This is a standard which defines 100Base-T operation over Category 3 or voice grade cabling which is not used for data communications at JPL)

Briefly and simplistically, this negotiation occurs by signaling over the connecting wire and is based on the signaling that is defined in the 10Base-T standard to establish "Link". Establishing link in a 10Base-T connection is done with a Network Link Test Pulse or NLP. 100Base-T devices use a Fast Link Pulse (FLP), which is a series of NLPs, when trying to negotiate with its link partner. The other side of the connection may not recognize FLPs and respond only with NLPs. This indicates to the 100Base-T device that its link partner is not "intelligent" and can only support 10Base-T. The 100Base-T device should fall back to 10Base-T operation if it supports it or not establish a connection (link) if it does not.

Auto-negotiation can be disabled and each end configured explicitly as to speed and dupelxing mode. This is typically referred to as "hard setting" the switch port or network interface card (NIC).

There are typically 3 components involved, two link partners and the cabling between the two.



?? The End station (usually a network device such as a workstation in an office or work area)

?? The Network Aggregation device (at JPL a network switch in a telecommunications closet)

?? The cabling which connects the two including the permanent cable run and the jumpers at either end used to complete the link

For additional details on how auto-negotiation works, several excellent Web resources are included at the end of this document.

Fast Ethernet (100Base-TX) support at JPL

JPLNet first introduced Fast Ethernet support as part of the HiNet task to cable the lab and install a hub (star) based network infrastructure. This first generation equipment only supported 100BaseTX Half Duplex. About 8 months later, JPLNet purchased a new model of switches which provided switched 10/100Base-T connectivity. This means that each port on the switch provides a dedicated link at:

10Mb/s	**Ethernet Notes**
10Base-T Half Duplex	Most common setting for older equipment and NICs
10Base-T Full Duplex	Not widely used at JPL. JPLNet recommends that if you think you need full duplex capability, you upgrade the NIC to one which supports 100Base-TX Full Duplex operation.
100Mb/s	**Fast Ethernet Notes**

100Base-TX Half Duplex	Phased out of JPLNet. Not recommended Unless your NIC only supports Half Duplex (some first generation Sun network interfaces only supported half duplex)
100Base-TX Full Duplex	**Recommended JPLNet setting.** This provides your system 100Mb/s in each direction and eliminates collisions.

The new generation of switches purchased supports 10/100 and some 1000 or Gigabit Ethernet ports for approved systems. This paper will focus on the 10/100 ports since this is the largest population and where the problems tend to manifest themselves.

1000Mb/s	**Gigabit Ethernet Notes**
1000Base-X Full Duplex	For approved systems with special requirements.

When JPLNet first introduced 10/100 capable switches, the Network Operations Center began to receive reports of problems as the original HiNet switches and hubs were upgraded to these new 10/100 capable switches. The problem was usually one of faulty negotiation, where each side had negotiated to different settings or not negotiated at all.

This problem manifests itself in a variety of ways. The most common symptoms seen at JPL are:

?? Problems with basic connectivity - No link / Switch port disables itself due to excessive errors (errDisable)

?? Poor network performance

?? Application problems

Problems with basic connectivity – No link / Switch port disables itself due to excessive errors (errDisable)

In some instances the negotiation was incompatible and connection was never achieved. As switches were upgraded at night, the next day users would come in, power up their systems, and find that they had no link to the network. Because many users turn their workstations off at night it is impossible for the JPLNet upgrade team to differentiate between, "system turned off" and "no link" on the new switches. When the affected users called in their problem, the Network Operations Center (NOC) staff would attempt to "hard" set the port to 10-half remotely from the NOC. In some instances this did not work and Network Operations had to put back the older 10Base-T switches to support these older network devices.

Poor network performance - Link achieved but mismatched duplexing mode resulting in poor performance

The more difficult situation resulted when link was established but the settings were different on each link partner. This usually occurred when both systems were operating at the same speed (both ends at 10 or both ends at 100) but at different duplex settings. This often occurs when devices support auto-sensing (of speed) but not auto-negotiation (of speed and duplexing mode). This is the most common problem. It is here that the JPLNet NOC staff must work with the end users to see which setting "works". Often this involves trying several setting and monitoring the switch port to see if the error counters increment.

In some instances, if systems were rebooted the NIC settings would be lost and this process would need to be duplicated.

Application Problems

Duplex mismatches have resulted in various applications problems. The NOC has seen instances where a duplexing mismatch has resulted in the loss of IPX connectivity or in NFS timeouts.

Traffic based on the UDP protocol is also very sensitive to this problem. Its connectionless, non-guaranteed delivery model is impacted because many data packets are "damaged" when there is a duplexing mismatch. Traffic based on TCP will recognized "bad" and missing data and will attempt to re-send the data.

While it seems a rather simple problem to solve, i.e. set both ends, many factors make it a complex one.

?? Auto-negotiation standard is optional

?? Vendor representation of NIC and switch capabilities are misleading

?? A wide variety of network interfaces are used at JPL

?? The components required to establish a successful connection fall under the control of several organizations

?? Users move connections

?? Users do not use provided station cables

?? Hardware/software dependent behavior

Auto-negotiation standard is optional

While a standard for how this negotiation is to take place has been defined, the Fast Ethernet specification 802.3u has made it optional. Also, many would state that the specification itself was not sufficiently explicit and has therefore contributed to these problems. In fact, JPLNet has seen auto-negotiation problem between different models of equipment from the same vendor. By policy and procedure all JPLNet switch to switch links are hard set.

Vendor representation of NIC and switch capabilities are misleading

Some vendors have implemented auto-sensing which only "negotiates" the speed of the link but not the duplexing mode. Some vendors call this "auto-negotiation" of speed which can be misleading and cause the buyer to think that the product supports true auto-negotiation.

While the switch equipment deployed at JPL is finite, there are a wide variety of Network Interface cards deployed at JPL. It has been difficult to establish a knowledge base defining the behavior of a particular card with a particular model switch. It is often difficult to determine NIC vendor, model number, and version, particularly for the average user.

The components required to establish a successful connection fall under the control of several organizations

Solving the problem takes the work of both JPLNet and a user familiar enough with the workstation and NIC capabilities to make the appropriate settings.

Recalling the 3 components involved in this link:

The End station (usually a network device such as a workstation)	This is usually provided by a user or by the OAO/DNS Alliance
The Aggregation unit (usually a network hub or switch)	This is provided by JPLNet
The cabling which connects the two	This is provided by JPLNet

While the JPLNet NOC staff can make the appropriate changes on the appropriate switch port, many times changes need to be made to the users' NIC. Oftentimes the user is not comfortable making these changes and so a local administrator or the OAO/DNS Alliance must be brought it.

Users Move Connections

When JPLNet completes a network connection request, JPLNet personnel scan the entire link from end to end. You may have noticed the barcodes on the grey network drops in your office, JPL network equipment is also bar-coded as are the racks which hold the equipment. When a new connection is made, the user work location, network drop and jack, hub room, rack, network equipment, and port to which the users' system is connected are scanned. This information is tied to the IP Address a user is given via the network connection request. When a system is moved to another port or when a system does not use the provided IP address, this link is broken and the information which JPLNet collects is invalidated. This makes it more difficult to track down a system and the port to which it is attached. It could also cause JPLNet to configure the wrong port on the switch, thus possibly impacting another user and not solving the problem for the original user.

In addition, when such a move is made by a user, the switch port now supporting the system may not be configured optimally for that system, thus starting the cycle again.

Users do not use provided station cables

The NOC has seen cases where users will use a station cable not provided by JPLNet. The system NIC supports auto-negotiation and link negotiates with the switch to 100BaseTX full-duplex, the fastest setting possible. However, 100Base-TX needs Category 5 cabling to operate, and a user used a Category 3 cable that happened to be available. JPLNet provides a Category 5 station cable with every network connection. It is important that this cable be used. As we move into Category 5 Enhanced and Category 6 cabling, it will become even more important that the cable delivered for a particular network connection be used for that connection.

Looking for problem signature

Hardware/software dependent behavior

As much as the NOC has tried to identify the specific behaviors of a particular component with a particular switch, the NOC's overriding experience is that each connection needs individual attention.

The NOC has even seen instances where both devices report the same settings via hard setting, and still had performance problems which were fixed by setting both ends to auto-negotiation.

Adding to the variables already discussed, vendors implement proprietary schemes for value added functionality when using their products end to end. When a link is comprised of multi-vendor components these proprietary schemes may cause unexpected results.

When the NOC suspects there is a negotiation problem, NOC personnel will look at specific error counters on the switch port supporting the workstation in question. Two key error counters that represent a clear signature of a duplexing mismatch are FCS and alignment errors. These types of errors may also indicate a problem with the station NIC card.

How to look for signatures of this behavior from the workstation end

Solaris

```
netstat -I
```

Host 1

```
Name  Mtu  Net/Dest      Address              Ipkts    Ierrs Opkts   Oerrs Collis Queue
le0   1400 137.78.171.0  rawlins.jpl.nasa.gov 13320557 0     1885522 2036  36068  1
```

Host 2

```
Name  Mtu  Net/Dest      Address                Ipkts    Ierrs Opkts   Oerrs Collis Queue
hme0  1500 137.78.167.0  net-dev-2.jpl.nasa.gov 32322303 6     2429414 0     0      0
```

Host 3

```
Name  Mtu  Net/Dest      Address                  Ipkts     Ierrs Opkts   Oerrs Collis Queue
ge0   1500 137.78.0.0    notclodsgig.jpl.nasa.gov 46084428 18     3929630 0     0      0
```

Note: Loopback information removed

Host 1 has a duplex mismatch.

Hosts 2 and 3 are operating in full duplex mode and therefore have no collisions.

A more detailed approach for Sun hme interfaces (use at your own risk)[1]

Note: The commands below must be run as root and all below examples assume a sun hme card

 # ndd -set /dev/hme instance 0 (this selects hme device instance 0 = hme0, instance 1 = hme1 (switch to instance you want to examine when multiple instances present i.e. 0-3 on quad ethernet card)

 # ndd -get /dev/hme transceiver_inuse (0=internal rj45 100baseTx connector, 1=external mii transeiver)

 # ndd -get /dev/hme link_status (0=down, 1=up)

 # ndd -get /dev/hme link_speed (0=10Mb, 1=100Mb)

 # ndd -get /dev/hme link_mode (0=half duplex, 1=full duplex)

How to check the capabilities the link partner (switch):

 # ndd -set /dev/hme instance 0 (or number of instance you want to examine)

 # ndd /dev/hme lp_autoneg_cap (0=link partner not have auto neg capability, 1= link partner has auto neg capability)

 # ndd /dev/hme lp_100fdx_cap (0=link partner does not support 100FDX 1=link partner supports 100FDX)

 # ndd /dev/hme lp_100hdx_cap (0=link partner does not support 100HDX 1=link partner supports 100HDX)

How to check the current hme settings:

ndd -get /dev/hme link_status
ndd -get /dev/hme link_speed
ndd -get /dev/hme link_mode

Results:
link_status (read only)
0 for Link Down
1 for Link up

link_speed (read only)
0 for 10 Mbps
1 for 100 Mbps

link_mode (read only)
0 for Half-Duplex mode
1 for Full-Duplex mode

<u>Supported Configurations for error free operation</u>

JPL supports two configurations:

Switch	**Host**	**JPLNet NOC Coordination required**
Auto (Default)	Auto (Default) `ndd -set /dev/hme adv_autoneg_cap 1` Verify:	No

	`ndd -get /dev/hme link_speed`
	`1`
	`ndd -get /dev/hme link_mode`
	`1`
	`ndd -get /dev/hme adv_100fdx_cap`
	`1`
	`ndd -get /dev/hme adv_autoneg_cap`
	`1`
	
	All other values should return 0:
	`ndd -get /dev/hme adv_100T4_cap`
	`0`
	`ndd -get /dev/hme adv_100hdx_cap`
	`0`
	`ndd -get /dev/hme adv_10fdx_cap`
	`0`
	`ndd -get /dev/hme adv_10hdx_cap`
	`0`	

Hard 100/ Full Duplex	Hard100/Full Duplex/Negotiation off
	
	`ndd -set /dev/hme instance 0`
	`ndd -set /dev/hme adv_100fdx_cap 1`
	`ndd -set /dev/hme adv_autoneg_cap 0`
	`ndd -set /dev/hme adv_100T4_cap 0`
	`ndd -set /dev/hme adv_100hdx_cap 0`
	`ndd -set /dev/hme adv_10fdx_cap 0`
	`ndd -set /dev/hme adv_10hdx_cap 0`
	
	Verify:
	
	`ndd -get /dev/hme link_speed`
	`1`
	`ndd -get /dev/hme link_mode`
	`1`
	`ndd -get /dev/hme adv_100fdx_cap`
	`1`
	`ndd -get /dev/hme adv_autoneg_cap`
	`0`
	
	All other values should return 0:
	`ndd -get /dev/hme adv_100T4_cap`
	`0`
	`ndd -get /dev/hme adv_100hdx_cap`
	`0`
	`ndd -get /dev/hme adv_10fdx_cap`
	`0`
	`ndd -get /dev/hme adv_10hdx_cap`
	`0`
		Yes

Note: other configurations are supported to achieve link. Many older systems, including older

SPARCs and systems with AUI transceivers often require a port to be hard set to 10/half duplex.

Switch ports will negotiate by default. JPLNet recommends that both link partners be configured to autonegotiate.

JPLNet also supports a "hard" configuration where the speed and duplex settings are "hard" set.

In either configuration it is imperative that the settings be configured on the host such that they persist across a reboot of the machine. If these settings are not permanent and the system is power cycled, it will come up in a miss-matched configuration and result in the problems described above. It is also imperative in the hardcoded case that you not move the host to a different port without coordination with the JPLNet NOC. If you hard set your hme interface and move the system to a port that has autonegotiation enabled, you will again have a miss-matched configuration. Whichever configuration you select we recommend setting the parameters in /etc/system.

Autonegotiate (Recommended)

Switch port set to autonegotiate

set local Sun to autonegotiate with the switch (link partner)
 # ndd -set /dev/hme adv_autoneg_cap 1
Verify with:
 # ndd -get /dev/hme adv_autoneg_cap
1

ndd -get /dev/hme link_speed
ndd -get /dev/hme link_mode
ndd -get /dev/hme adv_100fdx_cap
ndd -get /dev/hme adv_autoneg_cap

ndd -get /dev/hme adv_100T4_cap
ndd -get /dev/hme adv_100hdx_cap
ndd -get /dev/hme adv_10fdx_cap
ndd -get /dev/hme adv_10hdx_cap

Hardcoded Configuration

Host - Switch Configuration:

Switch: Port is hard set to 100 full-duplex (no negotiation)

Host:
Verify

ndd -get /dev/hme link_speed
1
ndd -get /dev/hme link_mode
1
ndd -get /dev/hme adv_100fdx_cap
1
ndd -get /dev/hme adv_autoneg_cap
0

All Other:
ndd -get /dev/hme adv_100T4_cap
0
ndd -get /dev/hme adv_100hdx_cap
0
ndd -get /dev/hme adv_10fdx_cap
0
ndd -get /dev/hme adv_10hdx_cap
0

Persistent Configuration:

Please try (if using /etc/rc2.d/S99...)

```
ndd -set /dev/hme instance 0
ndd -set /dev/hme adv_100T4_cap 0
ndd -set /dev/hme adv_100fdx_cap 1
ndd -set /dev/hme adv_100hdx_cap 0
ndd -set /dev/hme adv_10fdx_cap 0
ndd -set /dev/hme adv_10hdx_cap 0
ndd -set /dev/hme adv_autoneg_cap 0
```

or (if using /etc/system)

```
set hme:hme_adv_autoneg_cap=0
set hme:hme_adv_100T4_cap=0
set hme:hme_adv_100fdx_cap=1
set hme:hme_adv_100hdx_cap=0
set hme:hme_adv_10fdx_cap=0
set hme:hme_adv_10hdx_cap=0
```

Note that the order does make a difference.
The link is re-negotiated when the interface is ifconfig'ed up or when ndd ndd adv_autoneg_cap command is executed.

Windows

In a DOS window type `netstat –es`.

```
C:\WINDOWS>netstat -es
Interface Statistics

                           Received            Sent

Bytes                       1038815          105991
Unicast packets                 925            1067
Non-unicast packets            7326              58
Discards                          0               0
Errors                            0               0
Unknown protocols               992

IP Statistics

  Packets Received                   = 6199
  Received Header Errors             = 255
  Received Address Errors            = 285
  Datagrams Forwarded                = 0
  Unknown Protocols Received         = 0
  Received Packets Discarded         = 0
  Received Packets Delivered         = 5706
  Output Requests                    = 1146
  Routing Discards                   = 0
  Discarded Output Packets           = 0

  Output Packet No Route             = 0
  Reassembly Required                = 0
  Reassembly Successful              = 0
  Reassembly Failures                = 0
  Datagrams Successfully Fragmented  = 0
  Datagrams Failing Fragmentation    = 0
  Fragments Created                  = 0

ICMP Statistics

                            Received    Sent
  Messages                  184         171
  Errors                    0           0
  Destination Unreachable   0           2
  Time Exceeded             0           0
  Parameter Problems        0           0
  Source Quenchs            0           0
  Redirects                 0           0
  Echos                     0           166
  Echo Replies              149         0
  Timestamps                0           0
```

```
  Timestamp Replies         0           0
  Address Masks             0           0

  Address Mask Replies      0           0

TCP Statistics

  Active Opens                        = 145
  Passive Opens                       = 0
  Failed Connection Attempts          = 13
  Reset Connections                   = 0
  Current Connections                 = 0
  Segments Received                   = 767
  Segments Sent                       = 758
  Segments Retransmitted              = 141

UDP Statistics

  Datagrams Received    = 2274
  No Ports              = 2770
  Receive Errors        = 0
  Datagrams Sent        = 77
```

Macintosh

TBD

If you suspect you have a negotiation issue, please contact the JPL Network Operations Center at 4-8159 or via email at noc@jpl.nasa.gov.

NOTES and Additional Resources

University of New Hampshire Interoperability Labs – Tutorial

http://www.iol.unh.edu/training/

Fast Ethernet

http://www.iol.unh.edu/training/fe/aneg/index.html

Charles Spurgeon's Quick Reference Guide to Ethernet on the Web

http://www.ots.utexas.edu/ethernet/descript-100quickref.html

Auto Negotiation Chapter

http://www.ots.utexas.edu/ethernet/100quickref/ch13qr_1.html

Configuring Ethernet and Fast Ethernet Switching

http://www.cisco.com/univercd/cc/td/doc/product/lan/cat5000/rel_5_1/config/ether.htm

Understanding How Fast Ethernet Link Negotiation Works

http://www.cisco.com/univercd/cc/td/doc/product/lan/cat5000/rel_5_1/config/ether.htm#xtocid212136

Troubleshooting Ethernet

http://www.cisco.com/univercd/cc/td/doc/cisintwk/itg_v1/tr1904.htm

Notes on Cisco Equipment

http://www.cisco.com/univercd/cc/td/doc/product/lan/cat5000/rel_5_1/config/ether.htm

2924XL

Auto-sensing on each port detects the speed of the attached device and automatically configures the port for 10- or 100-Mbps operation, easing the deployment of the switch in mixed 10BaseT and 100BaseT environments.

Auto-negotiation on all 10/100 ports automatically selects half- or full-duplex transmission mode to optimize bandwidth.

http://www.cisco.com/univercd/cc/td/doc/product/lan/cat6000/relnotes/78_6218.htm

Module Troubleshooting 6509

This section contains troubleshooting guidelines for module problems:

?? If the Catalyst 6000 or 6500 family switch detects a port-duplex misconfiguration, the misconfigured switch port is disabled and placed in the "errdisable" state. The following syslog message is reported to the console indicating the misconfigured port has been disabled due to a late collision error.

```
SYS-3-PORT_COLL:Port 8/24 late collision (0) detected

%SYS-3-PORT_COLLDIS:Port 8/24 disabled due to collision

%PAGP-5-PORTFROMSTP:Port 8/24 left bridge port 8/24
```

Reconfigure the port-duplex setting and use the set port enable command to re-enable the port.

?? When you hot insert a module into a Catalyst 6000 or 6500 family chassis, be sure to use the ejector levers on the front of the module to seat the backplane pins properly. Inserting a module without using the ejector levers might cause the supervisor engine to display incorrect messages about the module. For module installation instructions, refer to the Catalyst 6000 Family Module Installation Guide.

?? Whenever you connect a Catalyst 6000 or 6500 family port that is set to auto-negotiate to an end station or another networking device, make sure that the other device is configured for auto-negotiation as well. If the other device is not set to auto-negotiate, the Catalyst 6000 or 6500 family switch auto-negotiating port will remain in half-duplex mode, which can cause a duplex mismatch resulting in packet loss, late collisions, and line errors on the link.

http://www.cisco.com/univercd/cc/td/doc/product/lan/cat6000/ios127xe/78_10404.htm

?? If you have an interface whose speed is set to auto connected to another interface whose speed is set to a fixed value, configure the interface whose speed is set to a fixed value for half duplex. Alternately, you can configure both interfaces to a fixed-value speed and full duplex.

?? Occasionally, host NICs connected to WS-X6248-TEL or WS-X6248-RJ-45 modules may incorrectly revert to half-duplex after an auto-negotiation failure. This problem is fixed in software release 5.2(3)CSX. (CSCdm88013)